Journal of Engineering Science and Technology
Special Issue on SIMASE, August (2021) 90 - 98
© School of Engineering, Taylor's University

LAYERED AND MODULARIZED ONTOLOGY ARCHITECTURE FOR INDONESIAN CULTURAL HERITAGE (INDOCH) BASED ON UNESCO CULTURAL HERITAGE TAXONOMY AND CIDOC CRM

WINANGSARI PRADANI[1,2,*],
ZAINAL A. HASIBUAN[3], HARRY BUDI SANTOSO[1]

[1]Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia
[2]Faculty of Science and Technology, Universitas Al Azhar Indonesia,
Jl. Sisingamangaraja, Kebayoran Baru, Jakarta Selatan 12110, Indonesia
[3] Faculty of Computer Science, Universitas Dian Nuswantoro,
Semarang, Indonesia
*Corresponding Author: winangsari@uai.ac.id

Abstract

Ontology has been widely used in data integration systems with its role as a global conceptualization medium for heterogeneous data sources. This research proposes three things; first is Layered and Modularized Ontology Development Integrated with Other Ontology (LMODIOO) which is a combination of several frameworks, methodologies, ontological artifacts and non-ontological artifacts to conceptualizing the IndoCH; second is the IndoCH ontology artifact generated through the application of LMODIOO; and third is the formalization and implementation of IndoCH. In the formalization there is a way to choose the appropriate CRM class for each of the 40 types of Indonesian CH descriptions. IndoCH ontology has been implemented in several types of interrelated CH objects, namely Marriage Customs, Culinary, Traditional Clothing, Oral Traditions and Performing Arts. The ontology is built, validated and implemented using Protégé and several python libraries such as RDFLib, OWLReady2, and nltk. The implementation of the IndoCH ontology on the type of CH mentioned above can answer competency questions, such that the knowledge about one specific type of CH, the relationship among the CH objects, and aggregation information of CH objects by ethnicities or by location.

Keywords: CRM, Cultural heritage ontology, Indonesian CH, Layered ontology, Modular ontology, UNESCO CH based taxonomy.

1. Introduction

Indonesia is the largest archipelagic country in the world with 17,508 islands, 3.1 million km2 of sea area, and 81,000 km of coastline. Indonesia is home to no less than 300 ethnic groups. These ethnic groups use 742 languages and dialects and carry out all cultural expressions and traditional customs [1]. Data on Indonesian CH objects spread across various parties. According to the government, there are 2319 cultural heritage institutions, 435 museums [2], and 2001 cultural communities [1]. These data are stored in various different formats, both digital and in the form of printed information. CH object data are also stored in many Indonesian individuals, in the form of CH objects, ways of life, and the art that are passed down from generation to generation. With 95% of 144 million people aged 15-50 who are connected to the internet every day [3], the potential for crowdsourcing of CH object data is huge.

Both crowdsourcing data and institutional data, if not integrated, will only circulate to a limited number of people and cannot be utilized optimally. This study proposes the development of an ontology of Indonesian cultural heritage that can be used to unified Indonesian CH data. The layered and modular ontology architecture was chosen as a form of ontology building for Indonesian CH because of its support for collaborative ontology development which ultimately guarantees the survival and usability of the ontology [4].

The IndoCH ontology artefact has 5 layers – foundation, core, top domain, bottom domain, and instances, adopted from ontology for software testing [5]. CIDOC CRM was chosen as the foundational and core ontology. CIDOC CRM is a widely known conceptual modeling language for documenting CH artifacts [6]. CRM has been an ISO standard since 2006 [6]. At the domain layers, concepts related to the types of CH are compiled based on the taxonomy of CH types established by UNESCO.

The conceptualization of IndoCH was built using the method proposed in this study, namely LMODIOO which is a combination of several frameworks, methodologies, ontological artifacts and non-ontological artifacts. The resulting ontology designs are implemented in the CH cluster which are related to each other, namely 7 modules of Marriage Customs, Culinary, Traditional Clothing, Oral Traditions of Folklore, Pantun, Music Performances, and Dance Performances. Ontology validation is carried out by fulfilling queries for 3 groups of competency questions, namely: (1) obtaining information about a particular type of CH object, for example about what is a traditional wedding ceremony, where and who are the actors present, (2) getting information about the relationship of the CH object with other types of CH objects, such as what cuisine is served at a traditional marriage ceremony, (3) information on the number of CH objects owned by certain ethnicities.

2. Materials and Methods

The PARCOURS ontology is a layered ontology that uses CIDOC CRM as the foundational layer. However, since CIDOC-CRM does not collect all the concepts needed to specifically describe the conservation-restoration field, the authors extend it by adding a middle layer. This layer is composed of the PARCOURS core ontology, which contains specific details of the restoration-conservation domain, and existing CRM sci extensions related to observation and measurement activities.

The bottom layer is populated with a set of domain-specific thesauri (controlled vocabulary terms) for the purposes of providing unambiguous terminology.

The BudayaKB Ontology [7] aims to bridge the gap between the CH text orientation information problem and a structured web such as Wikidata. The author builds 2 modules: the extraction module and the integration module. The extraction module extracts from Wikipedia to build BudayaKB, a KB about CH Indonesia. The integration module links CultureKB to Wikidata. The entities generated in the extraction step are then transformed into the RDF data model in the Generate KB step into KB Culture. The CH category was built using natural language processing techniques, namely the lexicon descriptor by adding entities manually from DBPedia, Wikidata, the e-Indonesiana portal, and words on the Indonesian Wikipedia.

Neon Methodology is an ontology development methodology that supports aspects of collaborative development, reuse of existing ontologies, and dynamic evolution of ontology networks. An ontology network is a collection of scattered ontologies that are connected to each other. Neon methodology creates and distinguishes 9 scenarios of ontology development related to the use of ontological and non-ontological sources. It is not only what sources are used, but also how they are used, whether they are used in whole, in part, by re-engineering or without processing. The scenarios are arranged starting from the first, namely the ontology which is made from beginning to end without using ontological or non-ontological sources. It goes through the whole process: specification, scheduling, conceptualization, formalization, implementation [8].

3. Results and Discussion

3.1. CRM class selection for tangible CH

Tangible object is a man-made object, which is associated directly with Physical Module. Structure Tangible is matched with E26 Physical Feature. Feature in CRM means sticking or embedding to other objects. Other tangible such as CH Buildings, Sites and Regions is paired with E18 Physical Thing which is a freestanding object.

CH Tangible Movable Artifacts, Antiques, Historical Objects, and Numismatics objects are paired with the CRM E22 Human-Made Object class which in its description includes items that can be moved or not. Type CH Specimen is paired with E20 Biological Object because it includes biological objects such as animals, plants, minerals, etc. The type of CH Manuscript is a combination of the E22 Human-Made Object, E33 Linguistic Object, and E34 Inscription classes. E33 was chosen because it relates to the content of the manuscript, and E34 because manuscript objects can stick to other objects (E24).

3.2. CRM class selection for Intangible 1: Oral traditions

Intangible-1 CH is closely related to language, namely oral traditions and regional languages. In CIDOC CRM, there is an E28 Conceptual Object class family which contains objects created by humans and human-generated data. With the principle of selecting the most appropriate class and being at the bottom of the hierarchical structure, the subclass description under E28 Conceptual Object is examined. Here is the E28 Conceptual Object class hierarchy (see Fig. 1).

E28 - - - - Conceptual Object
E90 - - - - - *Symbolic Object*
E73 - - - - - - - *Information Object*
E29 - - - - - - - - *Design or Procedure*
E31 - - - - - - - - *Document*
E32 - - - - - - - - - *Authority Document*
E33 - - - - - - *Linguistic Object*
E34 - - - - - - - - *Inscription*
E35 - - - - - - - *Title*
E36 - - - - - - - - *Visual Item*
E37 - - - - - - - - *Mark*
E34 - - - - - - - - - - *Inscription*
E41 - - - - - - *Appellation*
E42 - - - - - - - - *Identifier*
E35 - - - - - - - - *Title*
E95 - - - - - - - - *Spacetime Primitive*
E94 - - - - - - - - *Space Primitive*
E61 - - - - - - - - *Time Primitive*
E89 - - - - - Propositional Object

Fig. 1. Results of CRM class selection for intangible 1.

It has been revealed that there are 2 subclasses of E28 Conceptual Object, namely E89 Propositional Object and E90 Symbolic Object. The difference between the two is in E90 where the objects are symbols or a set of symbols, characters, identifiers, traffic signs, emblems, texts, data sets, images, musical scores, multimedia objects, computer program code or mathematical formulae, while E89 are stories, plots, procedural prescriptions, algorithms, laws of physics, images of topics. From these two classes, it is clear that the Type CH Oral Tradition is more towards the E89 Propositional Object.

Next, the E89 subclass was traced again to get the most appropriate class. Subclass E89 Propositional Object has 2 subclasses, namely E73 Information Object and E30 Right. Furthermore, E73 has 4 subclasses: E29 Design or Procedure, E31 Document, E33 Linguistic Object, E36 Visual Item. The descriptions and examples for E33 Linguistic Object are closest to the Intangible 1 Oral Traditions type. The following is a description of the E33 Linguistic Object: This class comprises identifiable expressions in natural language or languages.

3.3. CRM class selection for Intangible 2 performance arts

The Sound, Music, Dance, and Theatre CH types are paired with the CRM E28 Conceptual Object class. The description in class E29 states that the instances of this class are ideas or thoughts that are not objects. Subclass E28 in the form of sound (audio) and motion objects has not been written explicitly in CRM, only E36 which is a visual object.

3.4. CRM class selection for Intangible 3 social, rituals, and events

The CRM class for the Intangible 3 type is divided into 2, namely E7 Activity and E33 Linguistic Object. Types of CH Tradition, Folk Games, Folk Sports, and Traditional Economic Systems are included in the E7 Activity because they are

activities carried out by the community in a certain place at a certain time. The types of CH Customary Law, Social Organization System, and Traditional Kinship System are included in E33 because they are written in written form.

3.5. CRM class selection for Intangible CH Type 4: Traditional knowledge

The CH type Natural Knowledge, Cosmology, and Local Wisdom is paired with the E28 Conceptual Object because it is knowledge or the result of thought. The CH type of Traditional Medicine is at the core Conceptual Object but is more specific with the E29 Design or Procedure class because medicine usually has special procedures or steps that must be followed.

3.6. CRM class selection for Intangible CH Type 5: Craftmanship.

All types of CH Intangible 5 are a process of producing goods so that Culinary Making, Traditional Clothes Making, Fine Arts Making, Craft Making, Weapon Making, Technology Tools Making, Architecture Developing, and Transportation Manufacturing are paired with E29 Design or Procedure.

3.7. Membership for CH instances

Although each CH class has been mapped to certain CRM classes, in formalization the instances of this class are not directly members of that CRM class. Instead, a new class is created which is a subclass of the related CRM class. The reason for this treatment is so that the original CRM classes do not change their formalization, and also so that the subclasses can be more flexible, in the sense that they can be added to axioms and their properties or are combined with other CRM classes. For example, the CulinaryMaking class is a combination of the CRM E7_Activity and E29_Design_or_Procedure classes, so a new class is created (see Fig. 2):

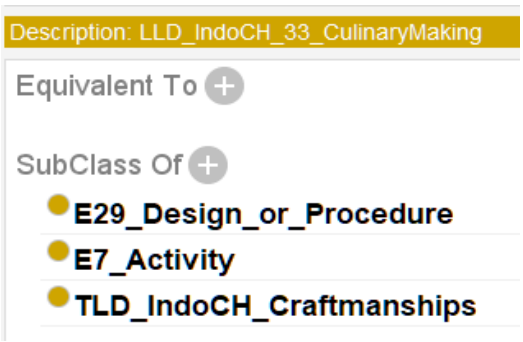


Fig. 2. Sample of CH instances.

3.8. Choosing properties

Selection of appropriate properties in CRM is also an effort that requires accuracy. The steps taken comply the following algorithm:

- Registering the attributes of the original data
- For each attribute, do:

- Look for the right property by paying attention to the domain class name and the appropriate range class name
- When the property is not found, create a new property name

For instance, Craftmanship has attributes of 'material', 'how to make', 'product name'.

In step 2a, the 'material' attribute corresponds to the P68_hasMaterial property, the 'product name' attribute corresponds to the P108_hasProduced property, while the 'how to make' attribute will be expressed as a literal. As in instance membership, a new property will be created as a subclass of the associated property, so the properties for the three attributes are presented in Table 1.

Table 1. Properties of the three attributes

Subject	Property	Object
TLD_IndoCH_Craftmanships	P68_IndoCH_Craftmanship_hasMaterial	E57_TLD_IndoCH_Craftmanship_NeededMaterial
TLD_IndoCH_Craftmanships	DP_IndoCH_TLD_Craftmanship_hasSteps	string
TLD_IndoCH_Craftmanships	P108_IndoCH_Craftmanship_hasProduced	E24_IndoCH_ProductOfCraftmanships

3.9. Ontology validation with SPARQL query

The following are queries that can be performed on the indoCH ontology:

Query Number of instances:

```
PREFIX coode: <http://www.co-ode.org/ontologies/ont.owl#>
PREFIX indoch: <D:…/Ontology/IndoCH#>
SELECT ?s
WHERE {
   ?s rdf:type coode:LLD_IndoCH_22_1_Wedding .
}
```

The query result is 572 LLD_IndoCH_22_1_Wedding instances.

Here is the number of instances of the other classes (see Fig. 3):

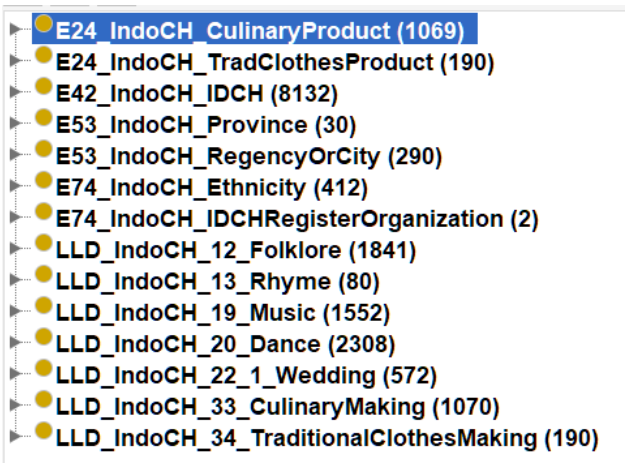


Fig. 3. Number of instances of other classes.

Query CH object on specific ethnicity

```
PREFIX coode: <http://www.co-ode.org/ontologies/ont.owl#>
PREFIX indoch:
<D://KANTOR/1_DIK16/0_SURVEY&OLAHDATA/Ontology/IndoCH#>
PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
SELECT ?ch
WHERE {
   ?ch a ?chSubClass .
   ?chSubClass rdfs:subClassOf* indoch:TLD_IndoCH_CulturalHeritage .
   ?ch coode:P14_IndoCH_isBelongToEthnicity indoch:Ethnicity.Jawa .
}
```

Query result: 807 CH object for Javanese ethnicity, among others are:

CH33_Cilok
CH33_Nyimplung
CH33_Gunung_Rembes
CH33_es_gronjong
CH12_Legenda_Dusun_Tapak
CH12_Legenda_Waduk_Naga_Wangsa
CH12_Rakyan_Petung
CH12_Legenda_Dusun_Mranak
CH12_Roro_Jongrang

```
PREFIX coode: <http://www.co-ode.org/ontologies/ont.owl#>
PREFIX indoch:
<D://KANTOR/1_DIK16/0_SURVEY&OLAHDATA/Ontology/IndoCH#>
SELECT ?culinaryMakingName
WHERE {
   ?culinaryMaking rdf:type indoch:LLD_IndoCH_33_CulinaryMaking .
   ?culinaryMaking coode:P7_TLD_CulturalHeritage_isInLocation ?dt2 .
   ?dt2 coode:P89_TLD_CulturalHeritage_isInProvince ?prov .
   ?prov rdfs:label "Papua" .
   ?culinaryMaking rdfs:label ?culinaryMakingName .
}
```

9
Culinary Making in Papua
1. Pembuatan Some
2. Pembuatan Dombo Teratai
3. Pembuatan Mesyedabu
4. Pembuatan Ufimba
5. Pembuatan Fumba
6. Pembuatan Pano
7. Pembuatan Akurec
8. Pembuatan Panarema
9. Pembuatan Mayu Ukwi

```
PREFIX cidoc: <http://www.cidoc-crm.org/cidoc-crm/>
PREFIX coode: <http://www.co-ode.org/ontologies/ont.owl#>
PREFIX indoch:
<D://KANTOR/1_DIK16/0_SURVEY&OLAHDATA/Ontology/IndoCH#>
SELECT ?adatPerkawinanName ?dt2_name ?prov_name
WHERE {
   ?s a coode:LLD_IndoCH_22_1_Wedding .
```

```
    ?s coode:P7_TLD_CulturalHeritage_isInLocation ?dt2 .
    ?s rdfs:label ?adatPerkawinanName .
    ?dt2 rdfs:label ?dt2_name .
    ?dt2 coode:P89_TLD_CulturalHeritage_isInProvince ?prov .
    ?prov rdfs:label ?prov_name .
  }
  LIMIT 10
```

Nakokne -> Malang, Jawa Timur
Megang -> Deli Serdang, Sumatera Utara
Ater Tolo -> Kota Pontianak, Kalimantan Barat
MELAMAR /NGITA NGUSOR KICIAN -> Ogan Komering Ulu Selatan, Sumat ra Selatan
Adat Moliyodu -> Gorontalo, Gorontalo
Mowa (pelepah pohon pinang) untuk löwölöwö zumange -> Nias, Sumatera Utara
saro-saro -> Kota Ternate, Maluku Utara
Mengeke -> Tapanuli Tengah, Sumatera Utara
Maminang -> Kota Sawahlunto, Sumatera Barat
Japuik Nikah -> Lima Puluh Kota, Sumatera Barat

4. Conclusions

In this research, an ontology of Indonesian cultural heritage (IndoCH) has been built with 5 layers of architecture, namely foundational, core, top level domain, low level domain and instances. Besides being layered, this cultural heritage ontology is also divided into modules based on function and structure. The ontology development methodology is a combination of several methodologies and guidelines: Generic Macro-Level Ontology Development [9], Neon Methodology (to create collaborative, dynamic, reused ontologies) [9], DSR-based layered ontology [5], Methodological Tips for Mappings to CIDOC-CRM [10]. The resulting ontology designs are implemented in the CH cluster which are related to each other, namely 7 modules of Marriage Customs, Culinary, Traditional Clothing, Oral Traditions of Folklore, Pantun, Music Performances, and Dance Performances. Several things that were found during the formalization of IndoCH were (1) property selection must be done carefully because some properties have the same purpose but for different domains and ranges; and (2) the addition of information can be done by adding entities without disturbing the original CRM related classes. Future work: This IndoCH can be developed according to the 9th scenario of Neon Methodology, which is to make IndoCH multilingual with at least 3 languages, namely Indonesian, English, and local languages from CH.

References

1. Kementerian Pariwisata. (2018). Peraturan menteri pariwisata nomor 13 tahun 2018 tentang rencana strategis kementerian pariwisata tahun 2018-2019. Retrieved January 15, 2022, from https://jdih.kemenparekraf.go.id/katalog-209-Peraturan%20Menteri.
2. Hadi, D.W.; Permanawiyat, W.; Sambodo, N.; Anindyatri, A.O.; and Mas' ad, M.A. (2019). *Statistik kebahasaan* 2019. Jakarta: Kementerian Pendidikan dan Kebudayaan.

3. Asosiasi Penyelenggara Jasa Internet Indonesia. (2020). Laporan survei internet APJII 2019-2020. Retrieved January 16, 2022, from https://apjii.or.id/ survei
4. Kotis, K.I.; Vouros, G.A.; and Spiliotopoulos, D. (2020). Ontology engineering methodologies for the evolution of living and reused ontologies: status, trends, findings and recommendations. *The Knowledge Engineering Review*, 35(e4), 1-34.
5. Olsina, L. (2021). Applicability of a foundational ontology to semantically enrich the core and domain ontologies. *Proceedings of the* 13*th International Conference on Knowledge Engineering and Ontology Development* (*KEOD*). Virtual Conference, Portugal, 111-119.
6. Meghini, C.; and Doerr, M. (2018). A first-order logic expression of the CIDOC conceptual reference model. *International Journal of Metadata Semantics and Ontologies*, 13(2), 131-149.
7. Putra, H.S.; Mahendra, R.; and Darari, F. (2019). BudayaKB: Extraction of cultural heritage entities from heterogeneous formats. *Proceedings of the* 9*th International Conference on Web Intelligence, Mining and Semantics*. Seoul Republic of Korea, 1-9.
8. Gómez-Pérez, A.; and Suárez-Figueroa, M.C. (2009). *NeOn methodology for building ontology networks: a scenario-based methodology*. Burgas: Demetra EOOD.
9. Keet, C.M. (2020). An introduction to ontology engineering. Retrieved January 16, 2022, from http://www.meteck.org/teaching/SA/OntoEngMeraka10ho.pdf
10. Theodoridou, M.; Bruseker, G.; Daskalaki, M.; and Doerr, M. (2016). Methodological tips for mappings to CIDOC CRM. *Proceedings of the* 44*th Conference on Computer Applications and Quantitative Methods in Archaeology*. Oslo, Norwegia, 89-102.